Localized Cumulative Distributions and a Multivariate Generalization of the Cramér–von Mises Distance

Uwe D. Hanebeck and Vesa Klumpp

***Abstract*— This paper is concerned with distances for comparing multivariate random vectors with a special focus on the case that at least one of the random vectors is of discrete type, i.e., assumes values from a discrete set only. The first contribution is a new type of characterization of multivariate random quantities, the so called Localized Cumulative Distribution (LCD) that, in contrast to the conventional definition of a cumulative distribution, is unique and symmetric. Based on the LCDs of the random vectors under consideration, the second contribution is the definition of generalized distance measures that are suitable for the multivariate case. These distances are used for both analysis and synthesis purposes. Analysis is concerned with assessing whether a given sample stems from a given continuous distribution. Synthesis is concerned with both density estimation, i.e., calculating a suitable continuous approximation of a given sample, and density discretization, i.e., approximation of a given continuous random vector by a discrete one.**

I. Introduction

A. Notation

$\underline{x}, \underline{\boldsymbol{x}}$	State vector and random vector $\in \mathbb{R}^N$
$f(\underline{x})$	Density function
$F^C(\underline{x})$	(Conventional) cumulative distribution
$F(\underline{x})$	Localized Cumulative Distribution (*LCD*)
$\delta(x)$	Scalar Dirac delta function
$\delta(\underline{x})$	Multidimensional Dirac delta function
$\Delta^C(\underline{x})$	(Conventional) cumulative distribution of Dirac delta function
$\Delta(\underline{x})$	LCD of Dirac delta density
$u(\underline{x})$	Uniform density
$U^C(\underline{x})$	Uniform cumulative distribution
$U(\underline{x})$	LCD of Uniform density
$N(., m, \sigma)$	Scalar Gaussian density
$N(., \underline{m}, \mathbf{C})$	Multidimensional Gaussian density
$\lvert\underline{x}\rvert$	Element wise absolute value of vector $\underline{x}$

B. Motivation

In many applications, it is necessary to compare two random vectors $\tilde{\underline{\boldsymbol{x}}}$, $\underline{\boldsymbol{x}}$ and their corresponding characterizations, i.e., their probability density functions $\tilde{f}(\underline{x})$ and $f(\underline{x})$. This includes both analysis and synthesis problems. In analysis, the closeness of two random vectors is investigated. Synthesis includes finding a random vector and its corresponding probability density function that is close to a given random vector and in some way simpler to handle or more convenient.

U. D. Hanebeck and V. Klumpp are with the Intelligent Sensor-Actuator-Systems Laboratory (ISAS), Institute of Computer Science and Engineering, Universität Karlsruhe (TH), Germany
uwe.hanebeck@ieee.org, klumpp@ira.uka.de

In many interesting cases, one or both random vectors considered are of discrete type on a continuous domain, i.e., assume values from a finite set only. In analysis, this leads to the famous statistical tests like the Kolmogorov–Smirnov test or the Cramér-von Mises test. In the case of one discrete random vector, these tests assess whether a given sample stems from a predefined continuous distribution. For two discrete random vectors it is determined whether their corresponding samples stem from the same underlying distribution.

In synthesis, we have to distinguish three cases:

- *One discrete and one continuous random vector:*
  - Case 1: Given the discrete random vector, e.g., samples, it is often desirable to perform an approximation by a continuous random vector and its associated density function. Applications include density estimation, interpolation, and parameter estimation.
  - Case 2: For a given continuous random vector and its associated density function, it has been found to be convenient to perform a systematic approximation by a discrete random vector [1], [2], [3], [4], [5] in order to simplify the processing steps required for Bayesian state estimation and filtering.
- *Two discrete random vectors:*
  - Case 3: Given a discrete random vector $\tilde{\underline{\boldsymbol{x}}}$, it is often useful to replace it by another discrete random vector $\underline{\boldsymbol{x}}$ with a different number of components. $\tilde{\underline{\boldsymbol{x}}}$ might represent redundant information, so that an approximation $\underline{\boldsymbol{x}}$ with a reduced number of components is sufficient. On the other hand, it might be required to enhance the resolution of $\tilde{\underline{\boldsymbol{x}}}$ by approximating it with $\underline{\boldsymbol{x}}$.

When at least one of the random vectors involved is of discrete type, comparing the two random vectors based on their probability density functions $\tilde{f}(\underline{x})$, $f(\underline{x})$ is impractical or even impossible. Hence, the corresponding cumulative distribution functions $\tilde{F}^c(\underline{x})$, $F^c(\underline{x})$ are employed for that purpose. In the case of a measured sample, the cumulative distribution is called the empirical distribution function (EDF).

The reverse problem of fitting the parameters of a continuous density function to a set of samples by minimizing

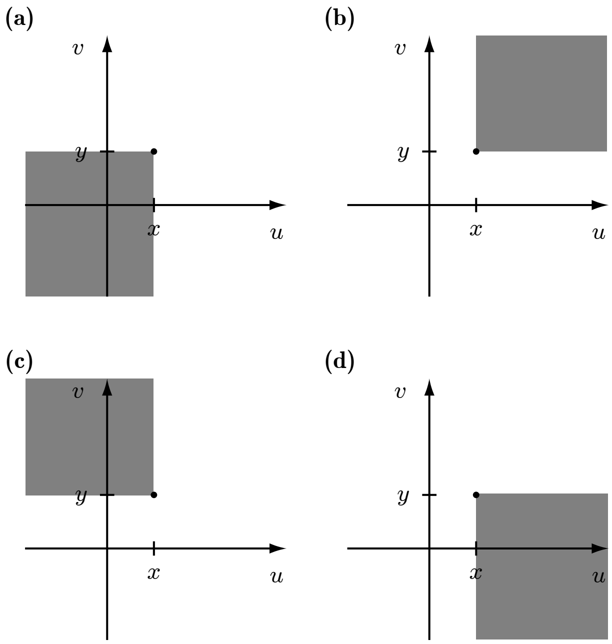

Fig. 1. Four different definitions of the conventional cumulative distribution in two–dimensional space. The value of the distribution at $[x, y]^T$ is evaluated by integrating the density function over the shaded area.

Kolmogorov–Smirnov test statistics is described in [6] for one–dimensional problems.

Employing the cumulative distributions works well in one dimension, i.e., for comparing random variables. However, in the multivariate case, cumulative distributions are not unique. For N–dimensional random vectors, there are 2^N different variants, $2^N - 1$ of which are independent.

Example I.1 (Cumulative Distributions in Two Dimensions) In two dimensions, there are four possible definitions of the cumulative distribution

1) $F_1(x, y) = P(\boldsymbol{x} \leq x, \boldsymbol{y} \leq y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u, v) dv du$ the standard definition, see Figure 1 (a),
2) $F_2(x, y) = P(\boldsymbol{x} > x, \boldsymbol{y} > y) = \int_{x}^{\infty} \int_{y}^{\infty} f(u, v) dv du$, Figure 1 (b),
3) $F_3(x, y) = P(\boldsymbol{x} \leq x, \boldsymbol{y} > y) = \int_{-\infty}^{x} \int_{y}^{\infty} f(u, v) dv du$, Figure 1 (c), and
4) $F_4(x, y) = P(\boldsymbol{x} > x, \boldsymbol{y} \leq y) = \int_{x}^{\infty} \int_{-\infty}^{y} f(u, v) dv du$, Figure 1 (d).

In addition, cumulative distributions are not symmetric, which is not a problem in one dimension. In several dimensions, however, this results in biases that depend on the selected variant.

The next example demonstrates these problems by means of the simple problem of approximating a two–dimensional uniform distribution with a single Dirac density with respect to a cumulative distance measure.

Example I.2 We consider a two–dimensional uniform density $u_{\underline{01}}(x, y)$ according to

$$u_{\underline{01}}(x, y) = u_{01}(x) \cdot u_{01}(y)$$

with

$$u_{01}(z) = \begin{cases} 1, & 0 \leq z \leq 1 \\ 0, & \text{elsewhere} \end{cases} ,$$

which is approximated by a Dirac mixture density

$$\delta(x - m_x, y - m_y) = \delta(x - m_x) \cdot \delta(y - m_y)$$

with

$$\delta(z) = \begin{cases} \text{undefined}, & z = 0 \\ 0, & \text{elsewhere} \end{cases} .$$

The corresponding conventional cumulative density distribution functions are given by

$$U_{\underline{01}}^C(x, y) = U_{01}^C(x) \cdot U_{01}^C(y)$$

with

$$U_{01}^C(z) = \begin{cases} 0, & z < 0 \\ z, & 0 \leq z \leq 1 \\ 1, & z > 1 \end{cases}$$

and

$$\Delta^C(x - m_x, y - m_y) = \Delta^C(x - m_x) \cdot \Delta^C(y - m_y)$$

with

$$\Delta^C(z) = \begin{cases} 0, & z < 0 \\ \frac{1}{2}, & z = 0 \\ 1, & z > 0 \end{cases} .$$

When applying the Cramér–von Mises distance [4] according to

$$D(m_x, m_y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} w(x, y) \left(U_{\underline{01}}(x, y) - \Delta^C(x - m_x, y - m_y) \right)^2 dx\, dy \ ,$$

we select $w(x, y)$ in such a way that only the region of interest is considered, e.g., $w(x, y) = u_{\underline{01}}(x, y)$. Hence, we obtain

$$D(m_x, m_y) = \int_0^1 \int_0^1 \left(x \cdot y - \Delta^C(x - m_x, y - m_y) \right)^2 dx\, dy \ .$$

The necessary condition for a minimum of $D(., .)$ gives

$$\begin{aligned} \frac{\partial D(m_x, m_y)}{\partial m_x} &= \int_0^1 \int_0^1 \left(xy - \Delta^C(x - m_x, y - m_y) \right) \delta(x - m_x) \Delta^C(y - m_y) dx\, dy \\ &= \int_0^1 \left(m_x y - \frac{1}{2} \Delta^C(y - m_y) \right) \Delta^C(y - m_y) dy \\ &= m_x \int_{m_y}^1 y dy - \frac{1}{2} \int_{m_y}^1 dy \\ &= m_x \ \frac{1}{2} y^2 \Big|_{m_y}^1 - \frac{1}{2} y \Big|_{m_y}^1 \\ &= \frac{m_x}{2} \left(1 - m_y^2\right) - \frac{1}{2} \left(1 - m_y\right) \\ &= -\frac{1}{2} \left(m_x m_y^2 - m_x - m_y + 1 \right) \stackrel{!}{=} 0 \end{aligned}$$

and in analogy

$$\frac{\partial D(m_x, m_y)}{\partial m_y} = -\frac{1}{2} \left(m_y m_x^2 - m_x - m_y + 1 \right) \stackrel{!}{=} 0 \ .$$

The solution

$$m_x = \frac{1}{2}(\sqrt{5} - 1), \ \ m_y = \frac{1}{2}(\sqrt{5} - 1)$$

does not correspond to the expected solution $m_x = \frac{1}{2}, m_y = \frac{1}{2}$ due to the missing uniqueness and symmetry of the conventional cumulative distribution.

As a result, the distance measures based upon the cumulative distribution are well suited and well established for one dimension only. In the multivariate case, several types of problems occur. The unweighted Cramér-von Mises distance for comparing two random vectors, for example, is zero when the two random vectors are identical and goes to infinity for the slightest difference between the corresponding densities. Hence, the definition of suitable distances for the multivariate case, i.e., between random vectors, is still an open problem.

Of course, some of these problems could be solved pragmatically. The non–uniqueness of the cumulative distributions could be treated by averaging the results obtained from all 2^N different variants, which unfortunately is impractical for high–dimensional random vectors. The unboundedness could be handled by either considering bounded domains as in the previous example or by using a suitable weighting function, i.e., by employing the weighted Cramér-von Mises distance, which might be impractical for densities with infinite support.

In summary, a definition of a modified type of cumulative distribution function that is suitable for multivariate random vectors would greatly enhance the applicability of the corresponding distance measures for comparing multivariate random quantities.

C. Prior Work

In general, two applications of a distance measure for probability distributions exist. These are the comparison of two distributions and the generation of a representation, i.e., samples, from a given distribution.

In order to determine the similarity between one–dimensional probability distributions, several approaches, like the Kolmogorov–Smirnov distance [7] or the Cramér–von Mises distance [8] can be used. Unfortunately, in general they are not suitable for multi–dimensional probability distributions. Determining the similarity between arbitrary multi–dimensional probability density distributions is still a mostly unsolved problem.

Testing the quality of samples regarding a given density function is referred to as *Goodness–of–Fit* test. Some approaches for the multi–dimensional case already exist. [10] describes the multi–dimensional Kolmogorov–Smirnov test and variations of it [11], [12], [13].

A new entropy measure based on the cumulative distribution of the considered random variable has been introduced in [9]. The so called *cumulative residual entropy* is well suited for both discrete and continuous random variables.

Comparing a Dirac mixture density [4] with an arbitrary density function, the individual Dirac components can be regarded as samples. Generating a deterministic sequence of samples from a given density distribution is often referred to as quasi–Monte Carlo sampling [14]. In the case of uniform distributions, so called low–discrepancy sequences [15] are used to determine regular approximations, unfortunately the best possible sequence is not known [16]. Different discrepancies for determining the quality of dispersion exist, e.g., L_2 discrepancy, [17] and star discrepancy [18]. These approaches are usually applied in numerical integration techniques [19] and the approximation of multivariate uniform distributions [20].

D. Key Ideas and Results of the Paper

In order to alleviate the problems of the conventional cumulative distribution, i.e., non–uniqueness and asymmetry, a different type of smoothing of the underlying probability density function is proposed. Rather than integrating over half–open infinite hyper spaces, an integration over finite domains is employed by defining certain window functions. The new characterization of the underlying probability density function, the so called Localized Cumulative Distribution, is then a function of the location of the window center and the window widths. Hence, for N–dimensional densities, the Localized Cumulative Distribution is $2 \cdot N$–dimensional.

In this paper, we restrict our attention to rectangular domains with arbitrary extents. This is similar to performing an rectangular wavelet transform [21] of the underlying probability density function, whereas there only the special case of a fixed number of rectangular domain sizes is considered. However, more general window types could be employed.

As the well–known distance measures between two random quantities based on cumulative distributions suffer from their associated problems especially in the multivariate case, a generalization of these distance measures can now be found by using the LCDs of the two random quantities to be compared. In this paper, a multivariate generalization of the well–known Cramér–von Mises distance is proposed by defining the squared integral deviation between the corresponding LCDs of the densities. This new type of distance measure can be used for the analysis and synthesis purposes described above.

Of course, the Localized Cumulative Distribution (LCD) can also be used for generalizing other types of distance measures like the Kolmogorov–Smirnov distance, i.e., the maximum or supremum over the difference of the LCDs corresponding to the two random quantities under investigation.

II. Localized Cumulative Distribution

Definition II.1 Let $\underline{\boldsymbol{x}}$ be a random vector with $\underline{\boldsymbol{x}} \in \mathbb{R}^N$, which is characterized by an N–dimensional probability density function $f : \mathbb{R}^N \to \mathbb{R}_+$. The corresponding Localized Cumulative Distribution (LCD) is defined as

$$F(\underline{x}, \underline{b}) = P\left(|\underline{\boldsymbol{x}} - \underline{x}| \leq \frac{1}{2}\,\underline{b}\right)$$

with $\underline{b} \in \mathbb{R}_+^N$ and $F(.,.) : \Omega \to [0, 1], \Omega \subset \mathbb{R}_+^N \times \mathbb{R}^N$.

Remark II.1 *Note that the relation $\underline{x} \leq \underline{y}$ with $\underline{x}, \underline{y} \in \mathbb{R}^N$ holds, if and only if $x_i \leq y_i$ for every element $i = 1 \ldots N$ of the vectors hold.*

Definition II.2 As a shorthand notation, we will denote the relation between the density $f(\underline{x})$ and its LCD $F(\underline{x}, \underline{b})$ by

$$f(\underline{x}) \circ\!\!-\!\!\bullet\, F(\underline{x}, \underline{b}) \ .$$

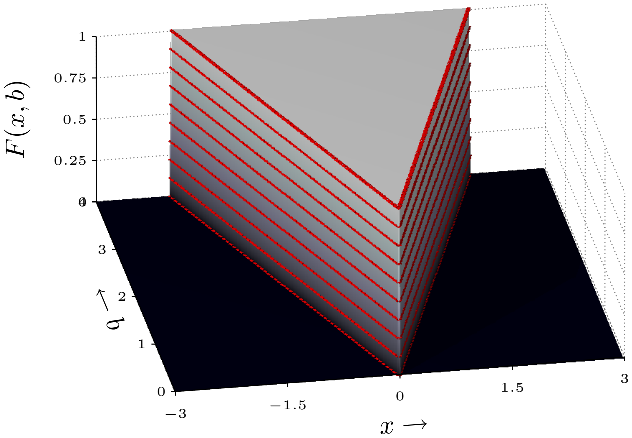

Fig. 2. Localized Cumulative Distribution of a Dirac density at location $x = 0$. The parameter b varies between 0 and 4.

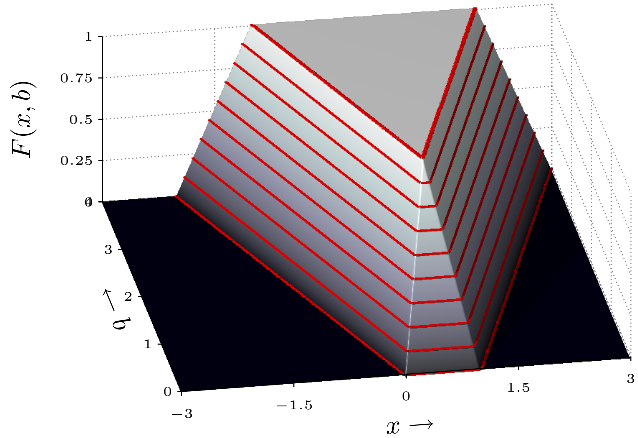

Fig. 3. Localized Cumulative Distribution of a one–dimensional uniform density over the interval $x \in [0, 1]$. The parameter b varies between 0 and 4.

Corrolary II.1 The Localized Cumulative Distribution function $F(\underline{x}, \underline{b})$ is calculated from the corresponding density function $f(\underline{x})$ according to

$$F(\underline{x}, \underline{b}) = \int_{\underline{x}-\frac{1}{2}\underline{b}}^{\underline{x}+\frac{1}{2}\underline{b}} f(\underline{t})\, d\underline{t} \quad .$$

PROOF. By definition, we have

$$F(\underline{x}, \underline{b}) = \int_{x_1-\frac{b_1}{2}}^{x_1+\frac{b_1}{2}} \int_{x_2-\frac{b_2}{2}}^{x_2+\frac{b_2}{2}} \cdots \int_{x_N-\frac{b_N}{2}}^{x_N+\frac{b_N}{2}} f_1(t_1) \cdot f_2(t_2) \cdot \ldots \cdot f_N(t_N)\, dt_N \ldots dt_2\, dt_1 \quad ,$$

which immediately gives the desired result. □

Theorem II.1 *The LCD $F(\underline{x}, \underline{b})$ has the following characteristics:*

1) $\forall i \in \{1, 2, \ldots, N\}$ *and finite* $\underline{b}_0$
$$\lim_{x_i \to \pm\infty} F(\underline{x}, \underline{b}_0) = 0 \quad ,$$
2) $\forall i \in \{1, 2, \ldots, N\}$ *and finite* $\underline{x}_0$
$$\lim_{b_i \to 0} F(\underline{x}_0, \underline{b}) = 0 \quad ,$$
3) *For finite* $\underline{x}_0$ *and* $\underline{\infty} = [\infty, \infty, \ldots, \infty]^T$
$$\lim_{\underline{b} \to \underline{\infty}} F(\underline{x}_0, \underline{b}) = 1 \quad .$$

Theorem II.2 (Separability) *For separable densities $f(\underline{x}) = f_1(x_1) \cdot f_2(x_2) \cdot \ldots \cdot f_N(x_N)$ with*

$$f_1(x_1) \circ\!\!-\!\!\bullet\, F_1(x_1, b_1) \quad ,$$
$$f_2(x_2) \circ\!\!-\!\!\bullet\, F_2(x_2, b_2) \quad ,$$
$$\vdots$$
$$f_N(x_N) \circ\!\!-\!\!\bullet\, F_N(x_N, b_N) \quad ,$$

the Localized Cumulative Distribution is also separable according to

$$f(\underline{x}) \circ\!\!-\!\!\bullet\, F(\underline{x}, \underline{b})$$

with

$$F(\underline{x}, \underline{b}) = F_1(x_1, b_1) \cdot F_2(x_2, b_2) \cdot \ldots \cdot F_N(x_N, b_N) \quad .$$

Theorem II.3 *The inverse transformation from a Localized Cumulative Distribution to its underlying density function is given as*

$$f(\underline{x}) = \lim_{b_1 \to 0} \cdots \lim_{b_N \to 0} \frac{F(\underline{x}, \underline{b})}{\prod_{i=1}^{N} b_i} \quad .$$

A. Special Case: One–dimensional Densities

In the case of a one–dimensional random variable $\boldsymbol{x}$ characterized by the density $f(x) : \mathbb{R} \to \mathbb{R}_+$, we have

$$F(x, b) = P\left(|\boldsymbol{x} - x| \leq \frac{b}{2}\right) = \int_{x-\frac{b}{2}}^{x+\frac{b}{2}} f(t) dt$$

with $F(x, b) : \Omega \to [0, 1]$, $\Omega \in \mathbb{R}_+ \times \mathbb{R}$.

In the scalar case, the Localized Cumulative Distribution can obviously be expressed in terms of the conventional cumulative distribution function according to

$$F(x, b) = F^c\left(x + \frac{b}{2}\right) - F^c\left(x - \frac{b}{2}\right) \quad .$$

Corrolary II.2 For separable densities $f(\underline{x})$ according to Theorem II.2 with individual LCDs given by

$$f_i(x_i) \circ\!\!-\!\!\bullet\, F_i(\underline{x}, \underline{b}) = F_i^c\left(x_i + \frac{b_i}{2}\right) - F_i^c\left(x_i - \frac{b_i}{2}\right)$$

for $i = 1, \ldots, N$, the total LCD is given by

$$f(\underline{x}) \circ\!\!-\!\!\bullet\, F(\underline{x}, \underline{b}) = \prod_{i=1}^{N} \left(F_i^c\left(x_i + \frac{b_i}{2}\right) - F_i^c\left(x_i - \frac{b_i}{2}\right)\right) \quad .$$

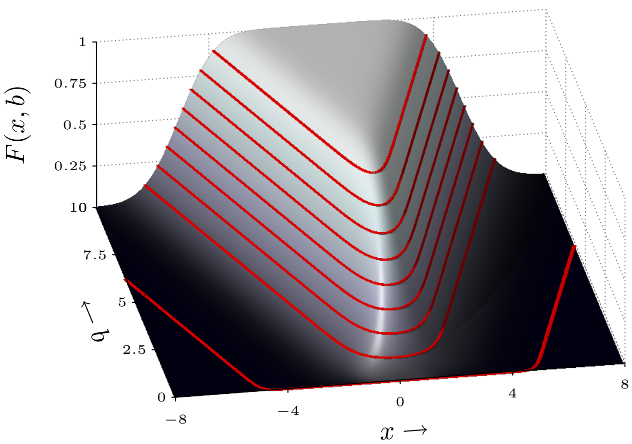

Fig. 4. Localized Cumulative Distribution of a one–dimensional Gaussian density with variance 1 and mean 0. The parameter b varies between 0 and 10.

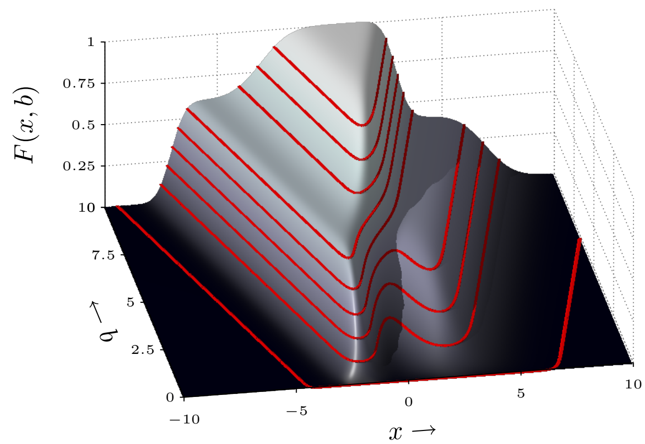

Fig. 5. Localized Cumulative Distribution of a one–dimensional Gaussian mixture density $f(x) = \frac{2}{5} N(x, 2, 1) + \frac{3}{5} N\left(x, -2, \frac{1}{2}\right)$. The parameter b varies between 0 and 10.

Example II.1 (Visualization of LCDs) The Localized Cumulative Distributions for various widths b are shown in Figure 2, Figure 3, Figure 4, and Figure 5. Figure 2 shows the LCD of a Dirac density at the location $x = 0$, i.e., $f(x) = \delta(x)$. In Figure 3 the LCD of the uniform density over the interval $[0, 1]$ is shown. Figure 4 visualizes the LCD of a Gaussian density $f(x) = N(x, 0, 1)$. Figure 5 visualizes the LCD of the non–symmetrical Gaussian mixture density $f(x) = \frac{2}{5} N(x, 2, 1) + \frac{3}{5} N\left(x, -2, \frac{1}{2}\right)$.

B. Special Case: Two–dimensional Densities

In the case of two–dimensional random vectors $\underline{\boldsymbol{x}}$ characterized by the density $f(x) : \mathbb{R}^2 \to \mathbb{R}_+$, we have

$$F(\underline{x}, \underline{b}) = P\left(|\underline{\boldsymbol{x}} - \underline{x}| \leq \frac{1}{2}\underline{b}\right) = \int_{x_1 - \frac{b_1}{2}}^{x_1 + \frac{b_1}{2}} \int_{x_2 - \frac{b_2}{2}}^{x_2 + \frac{b_2}{2}} f(t_1, t_2) dt_2 \, dt_1$$

with $F(x, b) : \Omega \to [0, 1]$, $\Omega = \mathbb{R}_+^2 \times \mathbb{R}^2$. It can be expressed in terms of the conventional cumulative distribution according to

$$\begin{aligned} F(\underline{x}, \underline{b}) = & F^C\left(x_1 + \frac{b_1}{2}, x_2 + \frac{b_2}{2}\right) + F^C\left(x_1 - \frac{b_1}{2}, x_2 - \frac{b_2}{2}\right) \\ & - F^C\left(x_1 - \frac{b_1}{2}, x_2 + \frac{b_2}{2}\right) - F^C\left(x_1 + \frac{b_1}{2}, x_2 - \frac{b_2}{2}\right) \quad . \end{aligned}$$

III. A Modified Cramér–von Mises Distance

Based on the definition of the Localized Cumulative Distribution (LCD), we will now present a modified version of the Cramér–von Mises distance suitable for comparing multivariate random quantities, i.e., random vectors.

The new distance is defined as the integral of the square of the difference between the LCD of the true density $\tilde{f}(\underline{x})$ and the LCD of its approximation $f(\underline{x})$.

Definition III.1 (Modified Cramér–von Mises Distance) The distance D between two densities $\tilde{f}(\underline{x}) : \mathbb{R}^N \to \mathbb{R}_+$ and $f(\underline{x}) : \mathbb{R}^N \to \mathbb{R}_+$ is given in terms of their corresponding LCDs $\tilde{F}(\underline{x}, \underline{b})$ and $F(\underline{x}, \underline{b})$ as

$$D = \int_{\mathbb{R}^N} \int_{\mathbb{R}_+^N} \left(\tilde{F}(\underline{x}, \underline{b}) - F(\underline{x}, \underline{b})\right)^2 d\underline{b} \, d\underline{x} \quad .$$

Remark III.1 *Obviously, the distance D has the following characteristics:*

1) *If the two densities $\tilde{f}(\underline{x})$ and $f(\underline{x})$ are equal, we have $D = 0$.*
2) *In the case of finite window LCDs, we have $D < \infty$ for arbitrary densities $\tilde{f}(\underline{x})$ and $f(\underline{x})$.*

Remark III.2 *A weighted variant D_w of the modified Cramér–von Mises Distance D is given by*

$$D_w = \int_{\mathbb{R}^N} \int_{\mathbb{R}_+^N} w(\underline{x}, \underline{b}) \left(\tilde{F}(\underline{x}, \underline{b}) - F(\underline{x}, \underline{b})\right)^2 d\underline{b} \, d\underline{x} \quad .$$

Remark III.3 *Of course, the Localized Cumulative Distribution can also be used for generalizing other cumulative distance measure such as the Kolmogorov–Smirnov distance.*

IV. Sanity Check

In order to demonstrate the usefulness of the new distance and especially its uniqueness and symmetry, we will reconsider a generalization of the problem posed in the Example I.2, i.e., we will approximate a multivariate uniform density by a single Dirac density.

Theorem IV.1 *For separable N–dimensional true densities $f(\underline{x})$ according to Theorem II.2, the locations of the optimal Dirac approximation that minimize the distance in Definition III.1 are obtained by solving*

$$\int_0^{b_i^{max}} \left(F_i^c(m_i - b_i) + F_i^c(m_i + b_i)\right) db_i = 2 \, b_i^{max} \, F_i^c(m_i)$$

for $i = 1, \ldots, N$.

PROOF. The distance measure is given by

$$D = \int_{\mathbb{R}_+^N} \int_{\mathbb{R}^N} \left(F(\underline{x}, \underline{b}) - \Delta(\underline{x} - \underline{m}, \underline{b}) \right)^2 \, d\underline{x} \, d\underline{b} \ ,$$

with $F(\underline{x}, \underline{b})$ according to Corrolary II.2 and

$$\Delta(\underline{x}-\underline{m}, \underline{b}) = \prod_{i=1}^{N} \left(\Delta^c\left(x_i - m_i + \frac{b_i}{2}\right) - \Delta^c\left(x_i - m_i - \frac{b_i}{2}\right) \right) \ .$$

The partial derivatives of D with respect to the location parameters m_j are given by

$$\frac{\partial D}{\partial m_j} = I_j^1 + I_j^2$$

for $i = 1, \ldots, N$, with the two expressions I_j^1 and I_j^2

$$I_j^1 = \int_{\mathbb{R}_+^N} \int_{\mathbb{R}^N} \Delta(\underline{x} - \underline{m}, \underline{b}) \, \frac{\partial \Delta(\underline{x} - \underline{m}, \underline{b})}{\partial m_j} \, d\underline{x} \, d\underline{b}$$

and

$$I_j^2 = -\int_{\mathbb{R}_+^N} \int_{\mathbb{R}^N} F(\underline{x}, \underline{b}) \, \frac{\partial \Delta(\underline{x} - \underline{m}, \underline{b})}{\partial m_j} \, d\underline{x} \, d\underline{b} \ .$$

The partial derivative $\frac{\partial \Delta(\underline{x}-\underline{m},\underline{b})}{\partial m_j}$ is given according to Lemma V.1 in the appendix. The first expression I_j^1 is zero according to Lemma V.2 in the appendix, so we only have to consider I_j^2 in the following derivation. For I_j^2, we obtain

$$I_j^2 = -\int_{\mathbb{R}_+^N} \int_{\mathbb{R}^N} \prod_{i=1}^{N} \left(F_i^c\left(x_i + \frac{b_i}{2}\right) - F_i^c\left(x_i - \frac{b_i}{2}\right) \right) \cdot \frac{\partial \Delta(\underline{x} - \underline{m}, \underline{b})}{\partial m_j} \, d\underline{x} \, d\underline{b} \ ,$$

which gives

$$\begin{aligned} I_j^2 = &-\int_{\mathbb{R}_+^N} \int_{\mathbb{R}^N} \left(F_j^c(x_j + \frac{b_j}{2}) - F_j^c\left(x_j - \frac{b_j}{2}\right) \right) \\ &\cdot \left(-\delta\left(x_j - m_j + \frac{b_j}{2}\right) + \delta\left(x_j - m_j - \frac{b_j}{2}\right) \right) \\ &\prod_{\substack{i=1 \\ i \neq j}}^{N} \left(F_i^c\left(x_i + \frac{b_i}{2}\right) - F_i^c\left(x_i - \frac{b_i}{2}\right) \right) \\ &\prod_{\substack{i=1 \\ i \neq j}}^{N} \left(\Delta^c\left(x_i - m_i + \frac{b_i}{2}\right) - \Delta^c\left(x_i - m_i - \frac{b_i}{2}\right) \right) \, d\underline{x} \, d\underline{b} \ . \end{aligned}$$

Using the properties of the Dirac delta function then gives

$$\begin{aligned} I_j^2 = &-\int_{\mathbb{R}_+} F_j^c(m_j - b_j) + F_j^c(m_j + b_j) - 2\, F_j^c(m_j) \, db_j \\ &\prod_{\substack{i=1 \\ i \neq j}}^{N} \int_{\mathbb{R}_+} \int_{\mathbb{R}} \left(F_i^c\left(x_i + \frac{b_i}{2}\right) - F_i^c\left(x_i - \frac{b_i}{2}\right) \right) \\ &\left(\Delta^c\left(x_i - m_i + \frac{b_i}{2}\right) - \Delta^c\left(x_i - m_i - \frac{b_i}{2}\right) \right) \, dx_i \, db_i \ . \end{aligned}$$

The necessary condition for a minimum is now obtained by $I_j^2 \stackrel{!}{=} 0$. For finite widths b_j, we finally obtain the desired result. □

Now, we will reconsider uniform densities as in Example I.2.

Example IV.1 For an N–dimensional uniform density with LCD $U(\underline{x}, \underline{b}) = \prod_{i=1}^{N} U_i(x_i, b_i)$ and

$$U_i(x_i, b_i) = U_i^c\left(x_i + \frac{b_i}{2}\right) - U_i^c\left(x_i - \frac{b_i}{2}\right) \ ,$$

applying Theorem IV.1 gives

$$\begin{aligned} \int_0^{b_i^{\max}} \left(U_i^c(m_i - b_i) + U_i^c(m_i + b_i) \right) \, db_i &= b_i^{\max} + m_i - \frac{1}{2} \\ &= 2\, b_i^{\max}\, U_i^c(m_i) \ . \end{aligned}$$

Dividing both sides by $b_i^{\max}$

$$1 + \frac{m_i - \frac{1}{2}}{b_i^{\max}} = 2\, U_i^c(m_i)$$

and taking the limit $b_i^{\max} \to \infty$ finally gives

$$U_i^c(m_i) \stackrel{!}{=} \frac{1}{2} \ ,$$

which provides the desired locations of the approximating N–dimensional Dirac density as

$$m_i \stackrel{!}{=} \frac{1}{2}$$

for $i = 1, \ldots, N$.

V. Discussion and Future Work

A generalization of cumulative distributions has been proposed, which can be interpreted as a cumulative rectangular kernel transform of the underlying probability density function. The new characterization of a random quantity, the so called Localized Cumulative Distribution (LCD) is unique and symmetric also in the multivariate case. By employing the LCDs of two random vectors rather than their conventional cumulative distributions, a multivariate generalization of the well–known Cramér–von Mises distance has been proposed. For the simple case of approximating a uniform density by a single Dirac density in arbitrary dimensions, it has been shown that the new distance measure gives useful results.

In the next step, the approach proposed in this paper will be used for generalizing the results obtained for approximating a continuous random variable by a discrete one (on a continuous domain) and vice versa. This includes the generalization of the density estimation proposed in [22] and the extension of the results derived in [1] for approximating a given continuous random variable and its associated density by a discrete random variable characterized by a Dirac mixture density.

Of course, the proposed Localized Cumulative Distribution could provide the basis for deriving new statistical tests similar to the Cramér–von Mises test or the Kolmogorov–Smirnov test, which are so far limited to the univariate case.

The proposed approach may also be useful for deriving alternatives to the current definition of entropy and mutual information valid for discrete or mixed discrete and continuous random vectors similar to the so called differential cumulative entropy proposed in [9].

References

[1] U. D. Hanebeck and O. C. Schrempf, "Greedy Algorithms for Dirac Mixture Approximation of Arbitrary Probability Density Functions," in *Proceedings of the 2007 IEEE Conference on Decision and Control (CDC 2007)*, New Orleans, Louisiana, Dec. 2007, pp. 3065–3071.

[2] O. C. Schrempf and U. D. Hanebeck, "Recursive Prediction of Stochastic Nonlinear Systems Based on Dirac Mixture Approximations," in *Proceedings of the 2007 American Control Conference (ACC 2007)*, New York, New York, July 2007, pp. 1768–1774.

[3] ——, "A State Estimator for Nonlinear Stochastic Systems Based on Dirac Mixture Approximations," in *Proceedings of the 4th International Conference on Informatics in Control, Automation and Robotics (ICINCO 2007)*, vol. SPSMC, Angers, France, May 2007, pp. 54–61.

[4] O. C. Schrempf, D. Brunn, and U. D. Hanebeck, "Density Approximation Based on Dirac Mixtures with Regard to Nonlinear Estimation and Filtering," in *Proceedings of the 2006 IEEE Conference on Decision and Control (CDC 2006)*, San Diego, California, Dec. 2006.

[5] ——, "Dirac Mixture Density Approximation Based on Minimization of the Weighted Cramér-von Mises Distance," in *Proceedings of the 2006 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI 2006)*, Heidelberg, Germany, Sept. 2006, pp. 512–517.

[6] M. D. Weber, L. M. Leemis, and R. K. Kincaid, "Minimum Kolmogorov-Smirnov Test Statistic Parameter Estimates," *Journal of Statistical Computation and Simulation*, vol. 76, no. 3, pp. 196–206, 2006.

[7] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, "Kolmogorov–Smirnov Test," in *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, 1992, pp. 623–628.

[8] D. D. Boos, "Minimum Distance Estimators for Location and Goodness of Fit," *Journal of the American Statistical association*, vol. 76, no. 375, pp. 663–670, 1981.

[9] M. Rao, Y. Chen, B. C. Vemuri, and F. Wang, "Cumulative Residual Entropy: A New Measure of Information," in *IEEE Transactions on Information Theory*, vol. 50, no. 6, June 2004, pp. 1220–1228.

[10] R. H. C. Lopes, I. Reid, and P. R. Hobson, "The two-dimensional Kolmogorov-Smirnov test," in *XI International Workshop on Advanced Computing and Analysis Techniques in Physics Research*, Amsterdam, The Netherlands, April 2007.

[11] J. A. Peacock, "Two-dimensional goodness-of-fit testing in astronomy," *Monthly Notices Royal Astronomical Society*, vol. 202, pp. 615–627, February 1982.

[12] G. Fassano and A. Franceschini, "A multidimensional version of the Kolmogorov-Smirnov test," *Monthly Notices Royal Astronomical Society*, vol. 225, pp. 155–170, 1987.

[13] A. Cooke, "The muac algorithm for solving 2d ks test." [Online]. Available: http://www.acooke.org/jara/muac/algorithm.html

[14] H. Niederreiter, *Random Number Generation and Quasi-Monte Carlo Methods*. SIAM, 1992.

[15] J. Hartinger and R. Kainhofer, *Monte Carlo and Quasi-Monte Carlo Methods 2004*. Springer, 2006, ch. Non-Uniform Low-Discrepancy Sequence Generation and Integration of Singular Integrands, pp. 163–179.

[16] J. Spanier and E. H. Maize, "Quasi-Random Methods for Estimating Integrals Using Relatively Small Samples," *SIAM Review*, vol. 36, no. 1, pp. 18–44, March 1994.

[17] W. J. Morokoff and R. E. Caflisch, "Quasi-Random Sequences and Their Discrepancies," *SIAM Journal on Scientific Computing*, vol. 15, no. 6, pp. 1251–1279, 1994.

[18] E. Thiémard, "An Algorithm to Compute Bounds for the Star Discrepancy," *Journal of Complexity*, vol. 17, no. 4, pp. 850–880, December 2001.

[19] J. Matoušek, "On the L2-Discrepancy for Anchored Boxes," *Journal of Complexity*, vol. 14, no. 4, pp. 527–556, December 1998.

[20] P. F. Hokayem and C. T. Abdallah, "Quasi-Monte Carlo Methods in Robust ControlDesign," in *Proceedings of the IEEE Conference on Decision and Control*, 2003, pp. 2435–2440.

[21] J. Benedetto and M. Frazier, Eds., *Wavelets: Mathematics and Applications*. CRC Press, 1994.

[22] A. Hanselmann, O. C. Schrempf, and U. D. Hanebeck, "Optimal Parametric Density Estimation by Minimizing an Analytic Distance Measure," in *Proceedings of the 10th International Conference on Information Fusion (Fusion 2007)*, Quebec, Canada, July 2007.

Appendix

Here, we will derive two lemmas useful for the derivations in Section IV.

The first Lemma is concerned with the partial derivative of the LCD corresponding to a multivariate Dirac density.

Lemma V.1 *The derivative of the LCD $\Delta(\underline{x}-\underline{m},\underline{b})$ corresponding to a multivariate Dirac density $\delta(\underline{x}-\underline{m})$ at location $\underline{m}$ with respect to the location parameter m_j in dimension j is given by*

$$\frac{\partial \Delta(\underline{x}-\underline{m},\underline{b})}{\partial\, m_j} = \left(-\delta\left(x_j-m_j+\frac{b_j}{2}\right)+\delta\left(x_j-m_j-\frac{b_j}{2}\right)\right) \prod_{\substack{i=1\\i\neq j}}^{N}\left(\Delta^c\left(x_i-m_i+\frac{b_i}{2}\right)-\Delta^c\left(x_i-m_i-\frac{b_i}{2}\right)\right) \; .$$

The second Lemma uses this result and provides a surprising result that proves useful for simplifying the derivations in Section IV.

Lemma V.2 *For the LCD of an N–dimensional Dirac density $f(\underline{x})=\delta(\underline{x}-\underline{m})$ at location $\underline{m}$ denoted by $F(\underline{x},\underline{b})=\Delta(\underline{x}-\underline{m},\underline{b})$, we have*

$$\int_{\mathbb{R}^N}\Delta(\underline{x}-\underline{m},\underline{b})\,\frac{\partial \Delta(\underline{x}-\underline{m},\underline{b})}{\partial\, m_j}\,d\underline{x}=0 \; ,$$

for $j=1,\ldots,N$, where m_j is the j–th component of $\underline{m}$.

Proof. Performing the integration gives

$$\begin{aligned}&\int_{\mathbb{R}^N}\Delta(\underline{x}-\underline{m},\underline{b})\,\frac{\partial \Delta(\underline{x}-\underline{m},\underline{b})}{\partial\, m_j}\,d\underline{x}\\ &=\int_{\mathbb{R}}\left(\Delta^c\left(x_j-m_j+\frac{b_j}{2}\right)-\Delta^c\left(x_j-m_j-\frac{b_j}{2}\right)\right)\\ &\quad\cdot\left(-\delta\left(x_j-m_j+\frac{b_j}{2}\right)+\delta\left(x_j-m_j-\frac{b_j}{2}\right)\right)\\ &\quad\prod_{\substack{i=1\\i\neq j}}^{N}\int_{\mathbb{R}}\left(\Delta^c\left(x_i-m_i+\frac{b_i}{2}\right)-\Delta^c\left(x_i-m_i-\frac{b_i}{2}\right)\right)\,dx_i\end{aligned}$$

For the first integral we obtain

$$\begin{aligned}&\int_{\mathbb{R}}\left(\Delta^c\left(x_j-m_j+\frac{b_j}{2}\right)-\Delta^c\left(x_j-m_j-\frac{b_j}{2}\right)\right)\\ &\quad\cdot\left(-\delta\left(x_j-m_j+\frac{b_j}{2}\right)+\delta\left(x_j-m_j-\frac{b_j}{2}\right)\right)\\ &=\Delta^c(-b_j)+\Delta^c(b_j)-2\,\Delta^c(0)=0 \; .\end{aligned}$$

Hence, the total derivative is zero, which concludes the proof. □